



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 6, 2023 – 08:23 PM EDT

PDB ID : 4FIP
Title : Structure of the SAGA Ubp8(S144N)/Sgf11(1-72, Delta-ZnF)/Sus1/Sgf73 DUB module
Authors : Samara, N.L.; Ringel, A.E.; Wolberger, C.
Deposited on : 2012-06-10
Resolution : 2.69 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.35
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.35

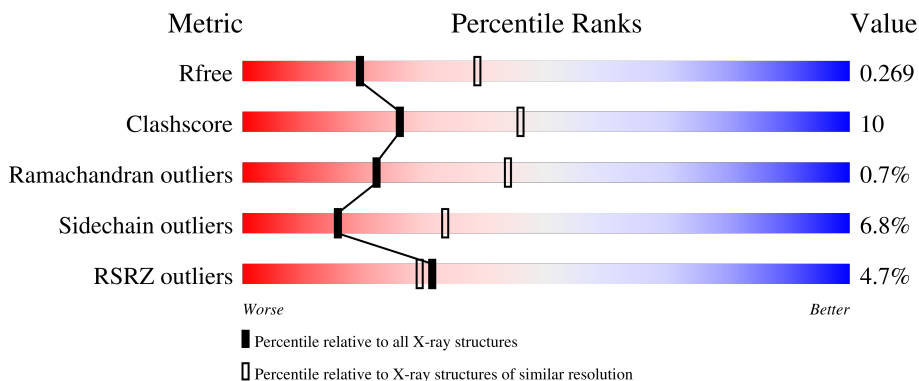
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.69 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



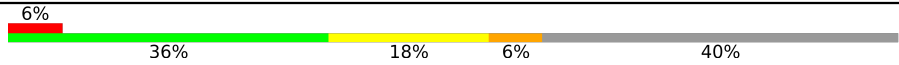
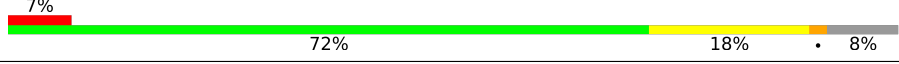

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3863 (2.70-2.66)
Clashscore	141614	4210 (2.70-2.66)
Ramachandran outliers	138981	4141 (2.70-2.66)
Sidechain outliers	138945	4141 (2.70-2.66)
RSRZ outliers	127900	3780 (2.70-2.66)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	476	 5% 74% 18% • 5%
1	E	476	 % 68% 23% • 7%
2	B	96	 5% 73% 18% • 6%
2	F	96	 4% 74% 16% • 6%
3	C	72	 3% 39% 14% • 43%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	72	
4	D	96	
4	H	96	

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 10730 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Ubiquitin carboxyl-terminal hydrolase 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	450	Total 3597	C 2276	N 618	O 667	S 36	0	0	0
1	E	441	Total 3534	C 2244	N 603	O 651	S 36	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP P50102
A	-3	ALA	-	expression tag	UNP P50102
A	-2	ALA	-	expression tag	UNP P50102
A	-1	ALA	-	expression tag	UNP P50102
A	0	ALA	-	expression tag	UNP P50102
A	144	ASN	SER	engineered mutation	UNP P50102
E	-4	GLY	-	expression tag	UNP P50102
E	-3	ALA	-	expression tag	UNP P50102
E	-2	ALA	-	expression tag	UNP P50102
E	-1	ALA	-	expression tag	UNP P50102
E	0	ALA	-	expression tag	UNP P50102
E	144	ASN	SER	engineered mutation	UNP P50102

- Molecule 2 is a protein called Protein SUS1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	90	Total 727	C 455	N 119	O 151	S 2	0	0	0
2	F	90	Total 735	C 460	N 122	O 151	S 2	0	1	0

- Molecule 3 is a protein called SAGA-associated factor 11.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	41	Total	C	N	O	0	0	0
			324	203	55	66			
3	G	43	Total	C	N	O	0	0	0
			342	213	57	72			

- Molecule 4 is a protein called SAGA-associated factor 73.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	88	Total	C	N	O	S	0	0	0
			695	439	115	137	4			
4	H	83	Total	C	N	O	S	1	0	0
			670	426	111	129	4			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	6	Total	Zn	0	0
			6	6		
5	D	1	Total	Zn	0	0
			1	1		
5	E	6	Total	Zn	0	0
			6	6		
5	H	1	Total	Zn	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	34	Total	O	0	0
			34	34		
6	B	3	Total	O	0	0
			3	3		
6	C	1	Total	O	0	0
			1	1		
6	D	11	Total	O	0	0
			11	11		
6	E	34	Total	O	0	0
			34	34		
6	F	5	Total	O	0	0
			5	5		
6	G	1	Total	O	0	0
			1	1		

Continued on next page...

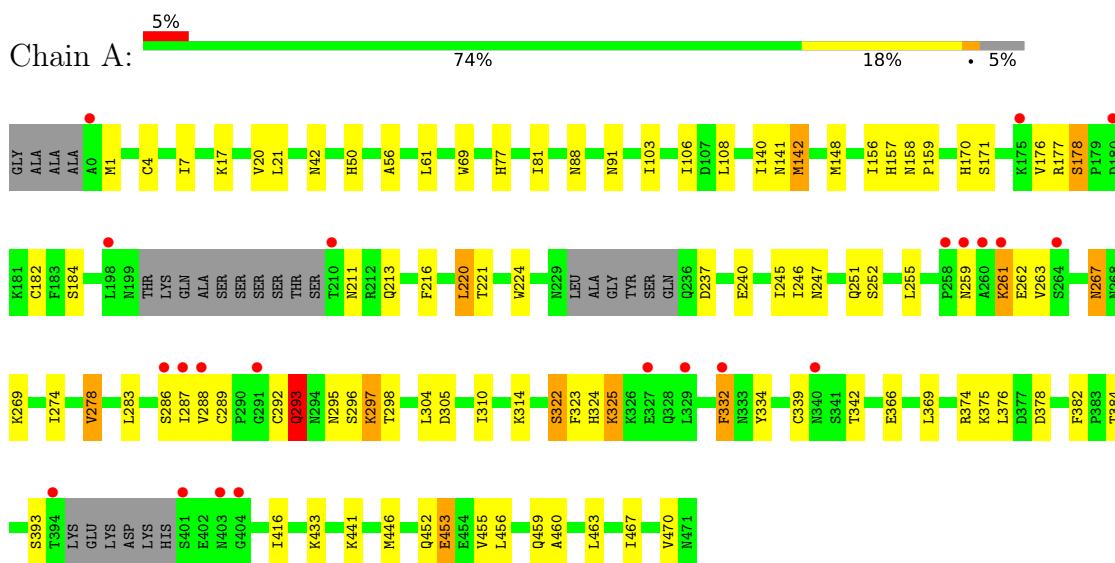
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	H	3	Total	O	0	0
			3	3		

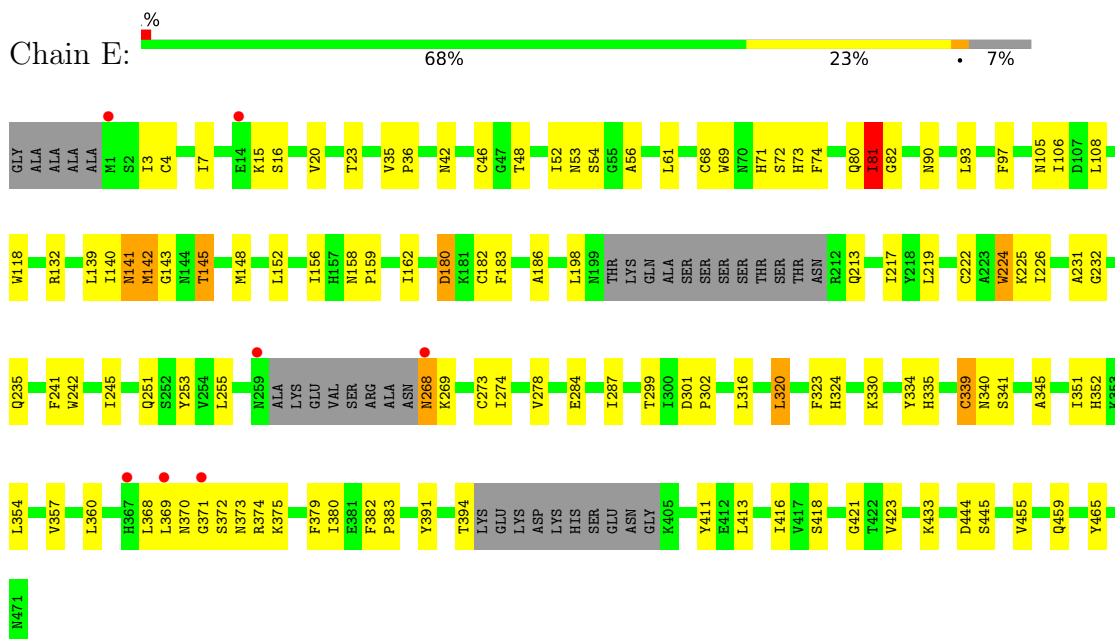
3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

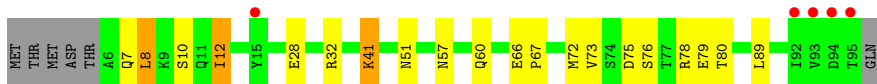
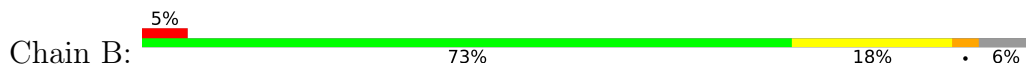
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 8



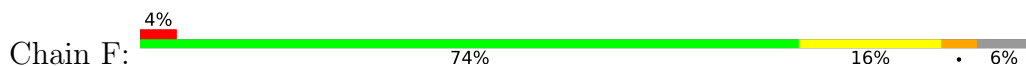
- Molecule 1: Ubiquitin carboxyl-terminal hydrolase 8



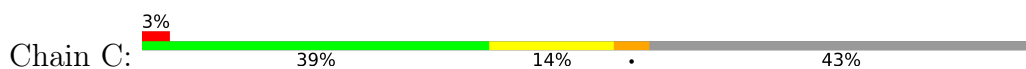
- Molecule 2: Protein SUS1



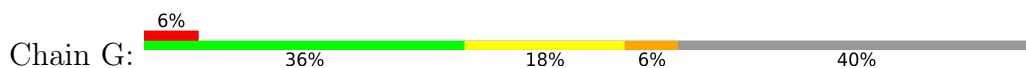
- Molecule 2: Protein SUS1



- Molecule 3: SAGA-associated factor 11



- Molecule 3: SAGA-associated factor 11

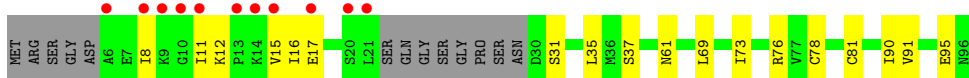


HIS

- Molecule 4: SAGA-associated factor 73



- Molecule 4: SAGA-associated factor 73



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	75.97Å 79.97Å 274.26Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.18 – 2.69 47.18 – 2.69	Depositor EDS
% Data completeness (in resolution range)	91.0 (47.18-2.69) 91.0 (47.18-2.69)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	0.15	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.72 (at 2.69Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.183 , 0.264 0.188 , 0.269	Depositor DCC
R_{free} test set	2199 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å ²)	45.3	Xtrriage
Anisotropy	0.601	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 42.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.025 for k,h,-l	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	10730	wwPDB-VP
Average B, all atoms (Å ²)	52.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.67% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section:
ZN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.71	1/3674 (0.0%)	0.78	1/4950 (0.0%)
1	E	0.68	2/3612 (0.1%)	0.77	2/4867 (0.0%)
2	B	0.62	0/733	0.75	0/988
2	F	0.62	0/744	0.75	0/1002
3	C	0.66	0/325	0.76	0/443
3	G	0.51	0/343	0.67	0/467
4	D	0.74	1/708 (0.1%)	0.83	0/954
4	H	0.67	0/683	0.76	0/919
All	All	0.68	4/10822 (0.0%)	0.77	3/14590 (0.0%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	69	TRP	CD2-CE2	5.43	1.47	1.41
1	E	224	TRP	CD2-CE2	5.15	1.47	1.41
4	D	32	TRP	CD2-CE2	5.14	1.47	1.41
1	E	69	TRP	CD2-CE2	5.13	1.47	1.41

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	237	ASP	CB-CG-OD1	5.79	123.51	118.30
1	E	81	ILE	N-CA-C	5.55	125.98	111.00
1	E	132	ARG	NE-CZ-NH1	5.52	123.06	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3597	0	3504	66	0
1	E	3534	0	3445	92	0
2	B	727	0	734	16	0
2	F	735	0	747	19	0
3	C	324	0	341	15	0
3	G	342	0	353	27	0
4	D	695	0	685	11	0
4	H	670	0	672	13	0
5	A	6	0	0	0	0
5	D	1	0	0	0	0
5	E	6	0	0	0	0
5	H	1	0	0	0	0
6	A	34	0	0	0	0
6	B	3	0	0	0	0
6	C	1	0	0	0	0
6	D	11	0	0	0	0
6	E	34	0	0	0	0
6	F	5	0	0	0	0
6	G	1	0	0	0	0
6	H	3	0	0	0	0
All	All	10730	0	10481	221	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 10.

The worst 5 of 221 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:89:LEU:HD11	3:G:11:ILE:HD11	1.20	1.19
1:E:105:ASN:HD22	3:G:40:THR:HG21	1.03	1.09
3:G:36:GLN:O	3:G:40:THR:HG23	1.60	1.02
2:B:12:ILE:HG22	4:D:11:ILE:HD11	1.47	0.96
1:A:251:GLN:O	1:A:255:LEU:HD13	1.65	0.95

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	442/476 (93%)	418 (95%)	22 (5%)	2 (0%)	29	52
1	E	433/476 (91%)	408 (94%)	22 (5%)	3 (1%)	22	44
2	B	88/96 (92%)	86 (98%)	2 (2%)	0	100	100
2	F	89/96 (93%)	81 (91%)	7 (8%)	1 (1%)	14	31
3	C	39/72 (54%)	38 (97%)	1 (3%)	0	100	100
3	G	41/72 (57%)	39 (95%)	1 (2%)	1 (2%)	6	13
4	D	82/96 (85%)	78 (95%)	3 (4%)	1 (1%)	13	29
4	H	79/96 (82%)	75 (95%)	3 (4%)	1 (1%)	12	27
All	All	1293/1480 (87%)	1223 (95%)	61 (5%)	9 (1%)	22	44

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	371	GLY
4	H	16	ILE
1	A	293	GLN
1	A	332	PHE
4	D	31	SER

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	410/429 (96%)	380 (93%)	30 (7%)	14	30

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	403/429 (94%)	384 (95%)	19 (5%)	26	50
2	B	85/91 (93%)	78 (92%)	7 (8%)	11	24
2	F	86/91 (94%)	77 (90%)	9 (10%)	7	14
3	C	39/68 (57%)	34 (87%)	5 (13%)	4	9
3	G	41/68 (60%)	37 (90%)	4 (10%)	8	17
4	D	79/86 (92%)	74 (94%)	5 (6%)	18	37
4	H	76/86 (88%)	72 (95%)	4 (5%)	22	45
All	All	1219/1348 (90%)	1136 (93%)	83 (7%)	16	33

5 of 83 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	320	LEU
2	F	44	ASP
1	E	330	LYS
1	E	459	GLN
3	G	8	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 44 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	141	ASN
1	E	348	GLN
1	E	170	HIS
1	E	276	HIS
2	F	13	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 14 ligands modelled in this entry, 14 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	450/476 (94%)	0.07	22 (4%) 29 27	25, 42, 90, 126	0
1	E	441/476 (92%)	-0.07	7 (1%) 72 73	29, 49, 79, 113	0
2	B	90/96 (93%)	0.11	5 (5%) 24 22	30, 50, 91, 104	0
2	F	90/96 (93%)	0.36	4 (4%) 34 32	37, 59, 104, 116	0
3	C	41/72 (56%)	0.05	2 (4%) 29 27	27, 38, 77, 91	0
3	G	43/72 (59%)	0.25	4 (9%) 8 6	40, 51, 89, 100	0
4	D	88/96 (91%)	0.38	7 (7%) 12 10	27, 48, 98, 101	0
4	H	83/96 (86%)	0.38	11 (13%) 3 2	31, 48, 105, 118	1 (1%)
All	All	1326/1480 (89%)	0.09	62 (4%) 31 29	25, 47, 92, 126	1 (0%)

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	95	THR	7.0
4	H	15	VAL	4.6
4	H	20	SER	4.4
4	H	13	PRO	4.2
2	B	95	THR	4.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
5	ZN	D	101	1/1	0.97	0.12	47,47,47,47	0
5	ZN	A	506	1/1	0.98	0.04	80,80,80,80	0
5	ZN	A	504	1/1	0.99	0.11	42,42,42,42	0
5	ZN	A	505	1/1	0.99	0.10	56,56,56,56	0
5	ZN	A	501	1/1	0.99	0.13	39,39,39,39	0
5	ZN	A	502	1/1	0.99	0.12	41,41,41,41	0
5	ZN	E	502	1/1	0.99	0.14	39,39,39,39	0
5	ZN	E	503	1/1	0.99	0.10	61,61,61,61	0
5	ZN	E	505	1/1	0.99	0.08	59,59,59,59	0
5	ZN	H	101	1/1	0.99	0.11	45,45,45,45	0
5	ZN	E	504	1/1	1.00	0.06	43,43,43,43	0
5	ZN	A	503	1/1	1.00	0.11	35,35,35,35	0
5	ZN	E	506	1/1	1.00	0.07	59,59,59,59	0
5	ZN	E	501	1/1	1.00	0.11	55,55,55,55	0

6.5 Other polymers [i](#)

There are no such residues in this entry.